Naval Research Laboratory

Monterey, CA 93943-5502

Three Dimensional Covariance Functions: Theory

RICHARD FRANKE

Atmospheric Dynamics & Prediction Branch
Marine Meteorology Division

DTIC QUALITY INSPECTED 2

October 1997

19971203 179

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. Agency Use Only *(Leave Blank)*.	2. Report Date. October 1997	3. Report Type and Dates Covered. Final

4. Title and Subtitle.
Three Dimensional Covariance Functions: Theory

5. Funding Numbers.
PE   0602435N
PN   BE 35-2-19

6. Author(s).
Richard Franke

7. Performing Organization Name(s) and Address(es).
Naval Research Laboratory
Marine Meteorology Division
Monterey, CA 93943-5502

8. Performing Organization Reporting Number.
NRL/MR/7531 - 97- - 7231

8. Sponsoring/Monitoring Agency Name(s) and Address(es).
Naval Research Laboratory (NRL)

10. Sponsoring/Monitoring Agency Report Number.

11. Supplementary Notes.

12a. Distribution /Availability Statement.
Approved for public release; distribution unlimited

12b. Distribution Code.

13. Abstract *(Maximum 200 words)*.

This report discusses some approaches to three dimensional covariance function modeling, suited for innovation data from a numerical weather prediction model. The incorporation of a method using a simultaneous domain transformation and fit to the covariance data appears to be a useful method of generating nonhomogeneous covariance functions. Some preliminary experience is discussed, along with plans for follow-on work.

14. Subject Terms.
Three Dimensional Covariance Functions: Theory

15. Number of Pages.
26

16. Price Code.

17. Security Classification of Report. UNCLASSIFIED	18. Security Classification of This Page. UNCLASSIFIED	19. Security Classification of Abstract. UNCLASSIFIED	20. Limitation of abstract. Same as report

NSN 7540-01-280-5500

i

Contents

1. Introduction

I begin by reviewing some relevant literature regarding covariance functions. A correlation function is a covariance function that has been normalized to have value one at the origin, and many properties are common. I will generally use the term covariance function unless there is a specific interest in having value one at the origin. Within the past dozen years or so there have been a number of relevant publications regarding the construction of covariance functions in two and three dimensions, tests for the requisite properties, and conditions on the parameters for certain functions.

In general I will not be specific about the precise mathematical properties various functions will need to possess, but suffice to say that if an expression involves an integral, for example, it is assumed the integral exists. A function is a valid covariance function provided it is positive definite (PD). A PD function can be characterized in several ways. Suppose C is a function defined on $\Re^d$. Then C is PD if (1) for any nonzero function g with finite support defined on $\Re^d$, $\int_{\Re^d} C(s-t)g(s)g(t)\,ds\,dt \geq 0$, or (2) for any finite set of points $s_i, i = 1,...,n$ and values $t_i, i = 1,...,n$, not all zero, the quadratic form $\sum_{i=1}^{n}\sum_{j=1}^{n} C(s_i - s_j)t_i t_j \geq 0$, (3) the Fourier transform of the isotropic function C(|s|) is nonnegative. In (1) and (2), if the equality can be attained the function C(s) is said to be positive semi-definite.

The last property is part of Bochner's Theorem (see, e.g., Priestly, [22]), which characterizes PD functions in that way. An excellent reference on PD functions is Stewart [27]. The set of isotropic covariance functions (in any dimension) can also be characterized as a probability mixture of Gaussian correlations (Matérn, [20]), as given by $\int_{-\infty}^{\infty} e^{-t^2 s^2}\,dF(t)$ where $F(t)$ is a one-dimensional cumulative probability distribution function. For certain choices of $F(t)$ one can easily obtain some families of covariance functions.

There are a number of simple building blocks for PD functions, as well as some easy tests to rule out certain functions. Convex combinations (linear combinations with

1

nonnegative coefficients summing to 1) of PD functions are PD. Products of PD functions are PD. Functions PD in $\Re^d$ are also PD in $\Re^D$ if $D < d$, indeed on any subset of $\Re^d$. Self-convolutions of functions are PD. There are certain inequalities on the values of correlation functions. A PD function is bounded by its value at the origin. A lower bound on the value of a correlation function, depending on the dimension of the space, also exists, e.g., $C(s) > -0.403..$ for d=2 and $C(s) > -0.219..$ for d=3.

Christakos [7] gives a review and a number of "criterion of permissibility" for checking whether a candidate function is indeed a covariance function. Franke [12] discusses some similar material, gives some ways of constructing PD functions and some tests for determining whether functions are PD, and also gives the results of actual and estimated errors resulting from the use of various correlation functions in some simulations. Much of this material also appears in Franke, et al. [14]. Thiébaux and Pedder [28] have some discussion of covariance models, and estimation of parameters. Daley [9] discusses the problem of estimating covariance functions. Watkins [29] considers the problem of estimating spatial covariance models, and in particular the problem of multimodality in maximum likelihood methods. Cressie [8] devotes considerable space to the problem of estimating the spatial relationships of data, generally approaching from the idea of using variograms rather than covariances, and mostly concentrating on geological and mining data. Weber and Talkner [30] discuss PD functions and investigate the question of whether certain functions proposed for use as covariance functions are valid, with some negative results. Gaspari and Cohn [17] discuss covariance functions and their properties and give some schemes for constructing covariance functions with special properties, such as finite support. Ron and Sun discuss positive definite function on spheres [20].

Almost all of the references in the previous paragraph concern isotropic and homogeneous functions. While this is an important class of functions, it is probably necessary to move beyond this class in numerical weather prediction applications. Some of the references discuss covariance functions on the sphere, clearly an important idea as objective analysis moves toward larger and larger volumes. From the fact that all functions that are PD in 3-space are also PD on subsets, this seems a simple problem.

2

However, most references note that when distance in 3-space is replaced by geodesic distance, the function may fail to be PD. In this case it is noted that geodesic distance $r\phi$, (ϕ is angle on the unit sphere) needs to be replaced by $2r\sin\dfrac{\phi}{2}$. It is generally not noted in the references that this returns the measure of distance to that in 3-space (i.e., through the interior of the sphere).

A recent new approach to nonhomogeneous and nonisotropic covariance modeling was proposed by Sampson and Guttorp [25]. While the precise manner in which the process was carried out, and later modified by Smith [26] leaves some problems to be solved, the basic idea seems rather attractive.

The question of how to generate nonhomogeneous covariance functions is complicated in general because the characterization of such functions is in terms of processes that are not easy to carry out. The characterization in terms of positive definiteness (PD-ness) of the function (or the discretization of it for some points) is mostly useful to show certain functions are not PD. The characterization in terms of the Fourier transform is for isotropic functions, and requires the calculation of integrals that may be impossible to carry out. In Sampson and Guttorp [25] the basic idea of a transformation of the data is brought forth (although there the transformation is described in terms of minimizing an error in the independent variables instead of the dependent variable). While Smith [26] is a bit unclear in his description of the details of his process, he is minimizing the error in the dependent variable. Both papers find the transformation for a set of discrete points, and complete the map using thin plate spline interpolation/ approximation. In what follows I will propose simpler processes with considerably less flexibility, and speculate about more flexible schemes.

The key idea that makes this a particularly interesting technique is that one can obtain nonhomogeneous, nonisotropic correlation functions, that is, functions that are guaranteed to be PD. The proof of PD-ness depends only on the fact that the transformation is one-to-one (thus barring repeated points after the transformation when none occurred in the original data) and that the transformed points are simply a set of points in the transformed space of the same dimension as the original set of points, hence

3

the correlation matrix must still be PD. The simplified process discussed in Section 4 can probably be implemented easily with significant potential benefit. For example, an existing procedure may be made to fit a problem better with only a transformation of variables.

2. The One Dimensional Case

In order to describe the ideas of the process easily, consider first the case of a set of time series of data depending on one spatial variable. Suppose we have measurements of the process at several points $x_1, x_2, ..., x_n$. Assume the x_i are in increasing order and that we have computed the correlation matrix with entries $C(|x_i - x_j|) = C_{i,j}$ for the data. We now want to fit a correlation function $F(d, \{a_k\})$ to approximate that data, where $\{a_k\}$ is a set of parameters to be determined, and d represents distance in a transformed variable with the transformation also being determined as part of the fitting process. Denote the monotonicity preserving transformation as $y = T(x)$, with the points $\{x_i\}$ being mapped to $\{y_i\}$. The variables that define the transformation may in fact be the values y_i, although the transformation could depend on fewer variables. Note that some care is needed to avoid dependency among the set of parameters $\{a_k\}$ defining the correlation function, and the set of parameters defining the transformation. The remainder of the process is simply to find those parameters by a minimization process. For example one might use least squares methods and minimize the objective function

$$\sum_{i=1}^{n} \sum_{j=1}^{i} \left[f(|x_i - x_j|, \{a_k\}) - C_{i,j} \right]^2 \tag{1}$$

over all parameters. This is a nonlinear problem, and the local minimum found by any minimization software will depend on the initial guess, emphasizing the importance of good initial guesses. Alternately the problem could be formulated to use a maximum likelihood method if desired.

An example of an application of this process to the data given by Lönnberg and Hollingsworth ([18] and [19]) is shown in Figures 1-4. The data used here is taken from [19] as estimated from their Figure 10, which represents the correlation of the error in

4

forecast heights at various pressure levels. Figure 1 shows the forecast error correlation data as a function of distance in the log(P) coordinate.
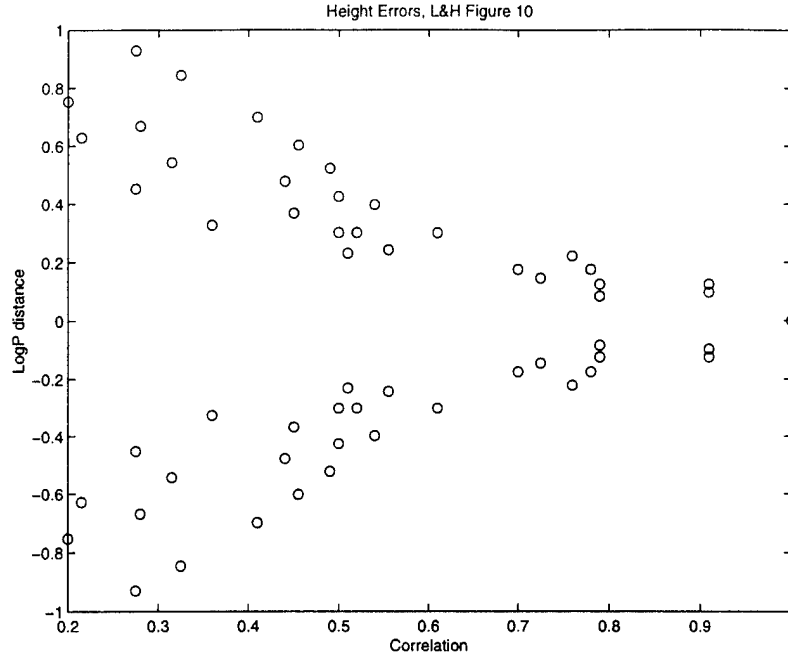

Figure 1: Height error correlation from Lönnberg and Hollingsworth [19].

To fit the data, a second order autoregressive correlation function (plus constant) of the form

$$F(d;a,c) = (1-c)\left(\cos ad + \frac{\sin ad}{a}\right)e^{-ad} + c \tag{2}$$

was used. Here d represents distance in the transformed variable, $T(x)$. Note that including the parameter in the exponential would cause a dependency in the absence of a constraint on the y_i because the transformation already embodies scaling. Constraining the $\{ y_i \}$ to lie in [0,1] (say) would lead to problems of how to impose the necessary constraints. If it seems desirable to have $\{ y_i \}$ in [0,1], that scaling can be carried out a posteriori. In order to enforce monotonicity of the y_i and eliminate a translation dependency, the parameters to be determined were taken to be $\{\delta y_i\}_{i=1}^n$ with $y_1 = 0$ and $y_{i+1} = y_i + |\delta y_i|$, i=1,...,n-1. Figure 2 shows the results of the calculation, including the

5

data and the fitted correlation function, now plotted as a function of distance in the transformed variable. Note how the transformation has removed much of the "scatter" from the data, especially at short distances.
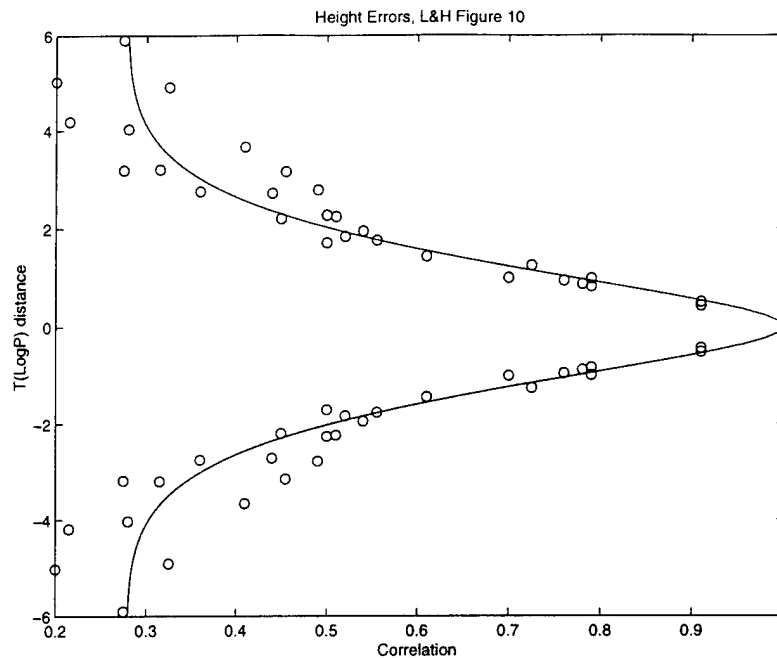

Figure 2: Height error correlation from Lönnberg and Hollingsworth [19] in the transformed coordinate, with second order AR fit.

When plotted back in the original log(P) distance in Figure 3, we see the nonhomogeneous, nonisotropic nature of the correlation functions (only four levels are plotted here of the eight given by the data). This example was computed using Matlab[1], using the function fmins to perform the optimization. To use the resulting correlation function in objective analyses, it will be necessary to determine the transformation of any point x to $y = T(x)$. The first inclination would be to use cubic splines. Cubic splines have continuous second derivatives and are straightforward to use.

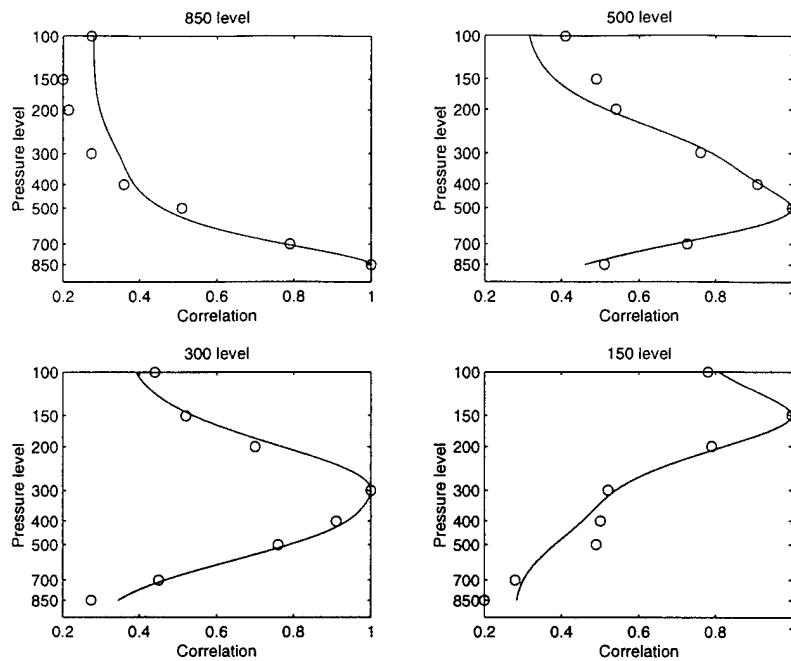[1] ® Registered trademark, The MathWorks, Natick, MA

Figure 3: The vertical correlation function for height errors for various pressure levels , with the data.

Figure 4 shows the spline function interpolating the points (x_i, y_i) that represents the transformation T. On the down side, however, cubic splines are not necessarily monotonic even when they are fit to data that have this property. If one continuous derivative of the transformation (and hence, the resulting correlation function) is sufficient, the monotonicity preserving piecewise cubic due to Fritsch and Carlson [16] will suffice. If higher order smoothness is necessary, exponential splines (splines in tension) can be used, although the selection of tension parameters is still something of an art. Software is available (Renka, [23]) that will attempt to iteratively determine the tension parameters to be as small as possible while still preserving monotonicty.
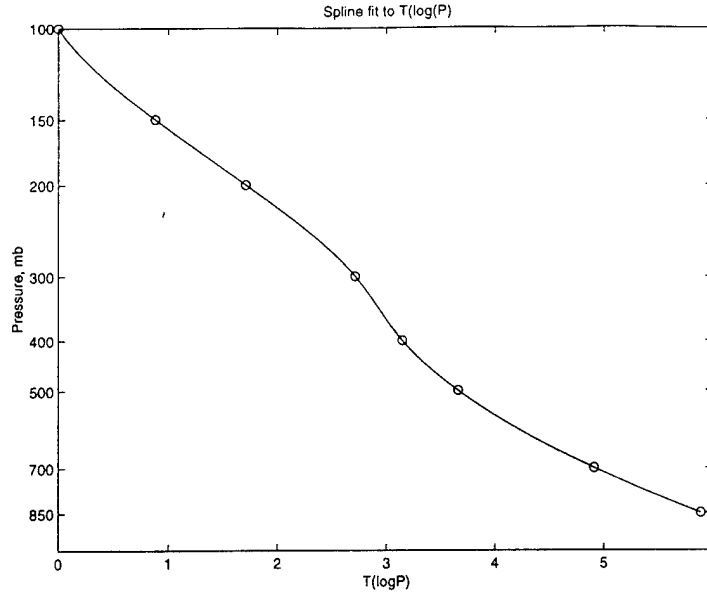
Figure 4: Spline fit to the T(logP) transformation.

Now, suppose the transformation is defined in such a way as to depend on fewer parameters than the n-1 used in the above example. It is as easily handled as above, and in the same manner. Let the points $\{X_j\}_{j=1}^m$ be prescribed in the interval $[x_1, x_n]$, and now let the transformation be $T : \{X_j\} \rightarrow \{Y_j\}$, with the parameters again being the increments δY_j with $Y_1 = 0$, and $Y_{j+1} = Y_j + |\delta Y_j|$, $j = 1,...,m-1$. The only difference from the previous application is that since calculation of the objective function requires $y_i = T(x_i)$, it is necessary to construct the interpolation function for each iteration of the optimization process.

3. The Two Dimensional Case

In the two dimensional case treated by Sampson and Guttorp [25], and by Smith [26], the transformation is defined by determining the points to which some subset of the data points are transformed, and then the map is completed by using thin plate spline interpolation/approximation. The details of exactly what has been done are sketchy. Needless to say, the imposition of one-to-one is difficult, and in fact it appears that the

8

map given by Smith in Figure 4.3 is not one-to-one. The procedure given here will guarantee a one-to-one map, while simplifying the transformation in one sense, although the software that calculates the transformation does require an iterative process.

The idea is to proceed in a manner strictly analogous to the one-dimensional process described in the previous section. Let the data locations be represented by $\{ \{(x_i, y_i)\}_{i=1}^{n}$. Assume that the model to be fit is $F(d, \{a_k\})$, and that d represents distance in the transformed space $T(x, y)$. The transformation $T(x, y)$ is determined by defining a tensor product monotonicity preserving interpolation function T. We could prescribe a grid $\{(X_j, Y_k)\}_{j=1, k=1}^{N, M}$ overlying the region of interest (one that contains all the points $\{(x_i, y_i)\}$). The parameters that are used in the optimization process will be $\{\delta u_j\}_{j=1}^{N-1}$ and $\{\delta v_k\}_{k=1}^{M-1}$ with $u_1 = v_1 = 0$ and $u_{j+1} = u_j + |\delta u_j|, j = 1, ..., N-1$ and $v_{j+1} = v_j + |\delta v_j|, j = 1, ..., M-1$.

The algorithm of Carlson and Fritsch [6] can be used to maintain monotonicity of the function that interpolates on the grid. If it is important to maintain continuity of the second derivative of the transformation, this technique will not work because the interpolant only has continuous first derivatives.

From this point the process is basically the same as described above. The expression to be minimized is the two dimensional version of Eq. (1), with the parameters in the minimization process being the $\{a_k\}$, the $\{\delta u_j\}$ and $\{\delta v_k\}$. As before, a maximum likelihood method formulation could be used if desired.

While this leads to a lot of control over the transformation, it also leads to merely stretching in each of the two variables and does not allow for any "twisting" type of behavior. Thus, while it would seem to be attractive, it does not embody enough flexibility. The problem of determining a transformation that contains parameters defined in such a way that the transformation will be one-to-one, and also has the flexibility to transform the set of grid rectangles to quadrilaterals with no overlap is unknown (to me) at this point. This is related to a process used in geography, imaging, biology, and other disciplines, called "warping"; a look at some of their research (e.g.,

[3], [5], [1], [11]) has not yet revealed whether they have considered the one-to-one problem associated with such transformations or whether they simply compute. Two additional references that do define some one-to-one transformation are Fitzpatrick and Leuze [10] where a polynomial transformation embodying nonlinear constraints on the parameters, and Franke and Hagen [15] where a one-to-one transformation using a biquadratic Bézier transformation with eleven free parameters is defined.

4. A Three Dimensional Case Under a Simpler Transformation

In meteorological problems, it may be sufficient to consider only the case of allowing parameters determining the horizontal correlation functions to vary with height. This would then be used in a product with the vertical correlation function to yield a partially separable correlation function of the form

$$C_P(s)C_h(d;\{a_k(P)\}) \tag{3}$$

where s is distance in the vertical variable, and d is distance in the horizontal variables. C_P is envisioned as being obtained under the kind of transformation discussed in Section 2, but here we concentrate on the correlation function for the horizontal plane, C_h. In theory as well as for practical purposes, it would be necessary for the function (3) to be PD in 3-space. To guarantee that the product would be PD in 3-space for arbitrary PD C_P, it is clear that C_h would have to be positive semi-definite in 3-space. It is possible that in practical problems, it might be permissible for C_h to be non-PD, because the product of C_P and C_h may still be PD, depending upon the function C_P. Numerical experiments have indicated that it may take a surprisingly small "nudge" from C_P to make the product with an non-PD C_h turn out to be PD. This will be mentioned in context later.

The problem of investigating the PD properties of (3) is simplified by requiring that C_h be PD. To investigate whether (3) is PD through the Fourier transform seems intractable. For many forms of C_h, the integrals over the horizontal plane are possible,

10

even straightforward. However, the parameter(s) $\{a_k\}$ then enter into the final integral over the vertical in complicated ways.

In applications, it will be necessary to have values of the parameters $\{a_k(P)\}$ for any P, not just at discrete levels. Further, if any of the parameters are not associated with a transformation of the particular P-plane, it is unclear how to enter those parameters into the calculation of C_h when two different P-values are involved. A simple example is illuminating. Consider the special form of the second order autoregressive function,

$$C_h(d,\{a(P),b(P)\}) = \left(\cos(a(P)d) + \frac{b(P)\sin(a(P)d)}{a(P)} \right) e^{-b(P)d} \tag{4}$$

with $a(P) = 0$. This yields $C_h(d;b(P)) = (1+b(P)d)e^{-b(P)d}$. Note that $b(P)$ depends on P, and d represents distance in the horizontal plane.

As written, this expression is ambiguous since, in general, two P-values are involved at the two points used to determine d. The obvious ploy of taking some average of $b(P)$-values does not necessarily result in a PD function. It is unknown at present whether there are conditions under which that might be possible.

For a known proper formulation it is necessary to go back to the basic representation $C_h(x_i,y_i,x_j,y_j;b) = (1+bd)e^{-bd}$, where again, $d^2 = (x_i - x_j)^2 + (y_i - y_j)^2$. We now think of first applying a transformation to the points (x_i,y_i,z_i) so that the transformed points are $(X_i,Y_i,z_i) = (b(P_i)x_i,b(P_i)y_i,z_i)$. Now, writing out the basic form of the function applied to the transformed points, we see the expression in terms of the original data is $C_H(x_i,y_j,x_i,y_j;b(P_i),b(P_j)) = (1+R)e^{-R}$ where

$R^2 = (b(P_i)x_i - b(P_j)x_j)^2 + (b(P_i)y_i - b(P_j)y_j)^2$. That is to say, $b(P)d$ is now replaced by distance in the transformed coordinates. To define $b(P)$ at all levels, a careful interpolation scheme that does not overshoot and is monotonic in intervals in which it should be monotonic can be used. An apparent potential problem is that R could be zero for two different points. However, this can only happen for points that have two different P-values (else the point is duplicated), and the vertical correlation function C_P

11

will prevent difficulty. Of course, in the case of a fixed value of b, the original data could have had repeated horizontal plane locations at different P values, so the phenomenon has nothing to do with the transformations considered here.

This scheme appears attractive because repeated fitting of innovation data for prediction height (say) at the various fixed levels would yield the parameter $b(P)$ and the estimates of observation and prediction error at each level. Subsequently fitting data at pairs of levels would yield the estimated values of the vertical observation error correlation and the vertical prediction error correlation (note that by the assumption of the form of the prediction error correlation made above, those would be the only quantities estimated when fitting data from two separate levels). After the transformation, the process is essentially that given by Hollingsworth and Lönnberg [18] with different horizontal correlation functions.

The additive constant used in the one-dimensional example has been found to be useful in practice (e.g., Barker [2]). An alternative that might achieve much of the same result but yields a correlation function that still goes to zero at large distances was proposed by Mitchell, et al. [21]. The idea was to add a second term of the same type but with a slower decay rate. A fixed ratio of the coefficients of the two terms was imposed. Either the additive constant or the latter idea would improve the fit. The possibility of allowing different coefficients for the additive term at different levels again raises the question of whether PD-ness would be maintained. As with $b(P)$, if the additive constant $c(P)$ were used, a careful interpolation would need to be used for intermediate levels. The question of how to handle the additive value for the correlation between levels is open. A few numerical experiments confirm that an average value (either geometric or arithmetic) seems to always result in a non-PD matrix. However, numerical experiments also indicate that multiplication by a C_P function, even one with a very slow decay rate is sufficient to correct the problem, so provided the additive constants do not vary too much, such problems may be insignificant.

The obvious next step is to attempt this process with the full generality of the second order autoregressive correlation function. Making this work through the transformation idea leads to difficulty as it seems not possible to work both parameters into the scheme

as transformations. Using a single value of the parameter a (after the transformation by $b(P)$ for all levels will give some additional flexibility and guarantee PD-ness.

We consider specifying the parameter a as $a(P)$. Again, it may be possible that certain restrictions on how the parameter $a(P)$ in equation (4) varies with P relative to $b(P)$ could result in a PD function in the context of the above. It is noted that Weber and Talkner [30] have shown that in the case when a and b are constants, the second order autoregressive function is PD only when $a^2 < 3b^2$. So, in general, (4) is not PD, but the allowable values of $a(P)$ in relation to $b(P)$ is considerably more complicated than in the case of fixed a and b, and will probably depend on the vertical correlation function C_h, as well. Another problem is in the precise way the parameters $a(P)$ in the denominator, and $b(P)$ multiplying the sine term are handled because of the two different P-values involved. Thinking in terms of transformations, and maintaining a scheme that reduces to the proper expression when the points are at the same P-value leads to

$$C_H(x_i, y_i, x_j, y_j, a(P_i), b(P_i), a(P_j), b(P_j)) = \left(\cos S + \frac{R \sin S}{S} \right) e^{-R},$$ where R is as given

above, and $S^2 = (a(P_i)x_i - a(P_j)x_j)^2 + (a(P_i)y_i - a(P_j)y_j)^2$.

Note that this brings a quotient of transformed distances into the expression to yield R/S which replaces the b/a value. Certain combinations of $a(P)$, $b(P)$, and observation locations do lead to non-PD matrices, and conditions on $a(P)$ and $b(P)$ to prevent this are not apparent.

The relationship between points on different P-values is dependent on the assumed coordinate system. A translation of the origin has no effect on a given level, however it will change the interlevel relationships. In particular, the relative location of points on different levels is altered by the translation. That is, the point that is transformed to the origin will affect the results.

This is very unsatisfying, although it might be possible to use it to advantage by setting the origin in some optimal way. One reasonable idea would be to transform the

13

center of the OI cell to the origin. A second alternative is to allow the parameters to vary with the two levels. It is easy to demonstrate that although in general this does not preserve PD-ness, when multiplied by a realistic vertical correlation function, the product may. To be definite about what is proposed we note the horizontal correlation function could be assumed to be of the form $C_h(d;a(P,Q)) = (1 + b(P,Q)d)e^{-b(P,Q)d}$, where P and Q are the pressure levels for the two points and d is the horizontal distance between them. This would be multiplied by the vertical correlation function as mentioned previously. Additional work is required to determine the practical facts.

5. Alternatives to Simple Transformations

The advantage of having a family of horizontal correlation functions that depend solely on a transformation varying with the P-level is compelling since that always guarantees PD-ness. On the other hand, the problem of building enough parameters into the transformation in order to adequately fit the data poses some difficulty. In the simplest extension, different transformation constants in longitude and latitude would yield an elliptical "equal distance" curve, which might be advantageous. Another possibility is to allow a rotation followed by an (independent) scaling in the variables, resulting in elliptical "equal distances" in the original variables, but now with the major axis of the ellipse at an angle that differs with P. Whether one would want to do this depends on how localized the correlation function calculation is to be. There would also be questions of how to properly interpolate for intermediate levels. A second alternative is to take the Lönnberg and Hollingsworth approach to modeling the correlation functions as a convex combination of correlation functions with height varying coefficients. While Lönnberg and Hollingsworth are fond of Bessel functions, one could use any other family, say the special second order autoregressive family with fixed parameters b_k for all levels. In this case the horizontal correlation function would have the form

$$C_h(d,\{A_k,b_k\}) = \sum_{k=1}^{K} A_k^2 (1 + b_k d)e^{-b_k d}$$ for which the A_k at each level would be found,

with the constraint $\sum_{k=1}^{K} A_k^2 = 1$. Here the A_k have been squared in the expression to emphasize that they must be positive. From the discrete height values at which the A_k are computed, one can now interpolate (carefully) for the values at other heights.

The above does not address the question of what form the horizontal correlation function in 3 coordinates should take. The suggestion by Lönnberg and Hollingsworth is not proper. Letting a second subscript on the A_k^2 denote a pressure level, they suggest the coefficients $\sqrt{A_{k,l}^2 A_{k,L}^2}$ for the coefficients of the correlation function between levels l and L. The sum of these coefficients, being geometric averages of coefficients that sum to one, is usually less than one, while an arithmetic average or weighted arithmetic average would preserve the constraint. It is unknown whether such a function would then be PD. Some experiments have been performed to investigate the matter empirically, with pessimistic results. Further experiments are necessary to see if multiplication by the vertical correlation function might restore PD-ness in realistic cases.

6. Miscellaneous

The scheme outlined by Hollingsworth and Lönnberg [18] for determining the vertical correlations of observation error and prediction error involve the assumption of certain correlation functions; in their case they were assumed to be the same for all levels. This avoids complications mentioned in the previous section. However, if different correlation functions are assumed for different levels, it is then necessary to assume some correlation function for data at two distinct levels. This was done by Lönnberg and Hollingsworth by assuming a particular kind of correlation function for the height thickness errors. When the height error correlation functions vary from level to level, the assumption of any form of correlation function for the thickness errors will lead to a complicated form for the relation between height errors at two different levels. Unless all work is performed in thickness errors instead of height errors, it seemed that it might be better to work solely in height errors. Subsequent work by the author [13] has demonstrated that the height-height errors have distinctly different behavior than that

exhibited by the second order autoregressive function, while the thickness-thickness errors are amenable to being fit by this function.

On further reflection, the Lönnberg and Hollingsworth idea of using the geometric mean of the coefficients probably arose from the process used to normalize a covariance matrix to a correlation matrix. In that process each entry is divided by the geometric mean of the diagonal element in its row and column. If they had done that, their scheme would have resulted in an assumed correlation function of the form that was the square root of the products of the correlation functions for the two levels involved. This seems attractive when one recalls that the product of a matrix and its transpose is PD;(unless singular). Unfortunately, the geometric mean is of the individual elements, not the matrix and its transpose, and such a matrix is not, in general, PD.

7. Modeling Real Data

An effort is presently underway to apply some of the ideas put forth here to model real data. The data is the innovation data (radiosonde observations minus forecast) for sixteen pressure height levels over most of North America (latitude $25°$ to $65°$ North, and longitude $70°$ to $130°$ West) for the October-November 1996 time period. The objective of the investigation is to determine a suitable correlation function structure that will enable the estimation of the forecast and observation errors, and in particular the vertical correlations. This effort could lead to more realistic modeling to be used during the objective analysis phase of the data assimilation cycle. The most important questions are the horizontal form of the covariance function, how the data is weighted in doing the approximations at various levels, and the assumptions about the covariance function for interlevel data. This requires investigating the data by looking at it in different ways (e.g., interlevel pressure height innovations, or thicknesses?) to try to determine what makes sense. In the case cited, assuming a form for one implies the form (a somewhat complicated form unless all horizontal levels use the same parameters) for the other. Some preliminary work has resulted in much of the software needed being now available. The goal of this work is to derive a vertically varying correlation model that satisfies the requisite PD-ness properties and incorporates an appropriate estimate of the vertically varying forecast and observation errors. The results of this study will be reported in [13].

16

References

1. Nur Arad and Daniel Reisfeld, Image warping using few anchors and radial functions, Comp. Graphics Forum 14(1994)35-46.

2. Edward H. Barker, Design of the Navy's multivariate optimum interpolation analysis system, Weath & Forecast. 7(1992)220-231.

3. I. Barrodale, D. Skea, M. Berkley, R. Kuwahara, and R. Poeckert, Warping digital images using thin-plate splines, Patt. Recog. 26(1993)375-376.

4. Peter Bartello and Herschel L. Mitchell, A continuous three-dimensional model of short-range forecast error covariances, Tellus 44A(1992)217-235.

5. Fred Bookstein, Principal warps: Thin-plate splines and the decomposition of deformations, IEEE Trans. Patt. Anal. and Mach. Intel. 11(1989)567-585.

6. Ralph E. Carlson and Fredrick N. Fritsch, An algorithm for monotone piecewise bicubic interplation, SIAM J. Numer. Anal. 26(1989)230-238.

7. George Christakos, On the problem of permissible covariance and variogram models, Water Resources Research 20(1984)251-165.

8. Noel A. C. Cressie, Statistics for Spatial Data, John Wiley & Sons, Inc., 1993.

9. Roger Daley, Atmospheric Data Analysis, Cambridge University Press, 1991.

10. J. Michael Fitzpatrick and Michael R. Leuze, A class of one-to-one two-dimensional transformations, Computer Vision, Graphic and Image Processing 30(1987)369-382.

11. David N. Fogel and Larry R. Tinney, Image registration using multiquadric functions, NCGIA TR#96-01, UCSB, 1996.

12. Richard Franke, Covariance functions for statistical interpolation, Naval Postgraduate School TR# NPS-53-86-007, September 1986.

13. Richard Franke, Three dimensional covariance functions: Real data, Naval Research Laboratory/Monterey Memorandum Report #NRL/MR/7531-97-7232, October 1997.

14. Richard Franke, Edward Barker, and James Goerss, The use of observed data for the initial value problem in numerical weather prediction, Computers and Mathematics with Applications 18(1988)169-184.

15. Richard Franke and Hans Hagen, Least squares surface approximation using multiquadrics and parametric domain distortion, submitted to Computer Aided Geometric Design, 25 September 1997.

16. Fredrick N. Fritsch and Ralph E. Carlson, Monotone piecewise cubic interpolation, SIAM J. Numer. Anal. 17(1980)238-246.

17. Gregory Gaspari and Stephen E. Cohn, Construction of correlation functions in two and three dimensions, to appear in the Quarterly Journal of the Royal Meteorological Society.

18. A. Hollingsworth and P. Lönnberg, The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, Tellus 38A(1986)111-136.

19. P. Lönnberg and A. Hollingsworth, The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors, Tellus 38A(1986)137-161.

19. P. Lönnberg and A. Hollingsworth, The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors, Tellus 38A(1986)137-161.

20. B. Mátern, Spatial Variation, second edition, Lecture Notes in Statistics, No. 36, Springer, 1986.

21. Herschel L. Mitchell, Cécilien Charette, Clément Chouinard, and Bruce Brasnett, Revised interpolation statistics for the Canadian data assimilation procedure: their derivation and application, MWR 118(1990)1591-1614.

22. M. B. Priestley, Spectral Analysis and Time Series, Vol. 1: Univariate Series. and Vol. 2: Multivariate Series, Prediction and Control, Academic Press, 1981.

23. Robert L. Renka, Algorithm 716 TSPACK: Tension spline curve-fitting package, TOMS 19(1993)81-94.

24. Amos Ron and Xingping Sun, Strictly positive definite functions on spheres in Euclidean spaces, Math. Comp. 65(1996)1513-1530.

25. Paul D. Sampson and Peter Guttorp, Nonparametric estimation of nonstationary spatial covariance structure, JASA 87(1992)108-119.

26. Richard L. Smith, Estimating nonstationary spatial correlations, manuscript, June 13, 1996.

27. J. Stewart, Positive definite functions and generalization, an historical survey, Rocky Mountain Journal of Mathematics 6(1976)409-434.

28. H. J. Thiébaux and M. A. Pedder, Spatial Objective Analysis, Academic Press, 1987.

29. A. J. Watkins, On models of spatial covariance, Comp. Stat. & Data Analysis 13(1992)473-481.

30. Rudolf O. Weber and Peter Talkner, Some remarks on spatial correlation function models, MWR 121(1993)2611-2617.